REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-01-

0315

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		01 Jul 97 to 31 Aug 00 Final

4. TITLE AND SUBTITLE
NDR Devices and Their Digital Applications

5. FUNDING NUMBERS
61103E
C757/14

6. AUTHOR(S)
Dr Mazumder

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan
Dept of Electrical Engineering & Computer Science
1301 Beal Avenue
Ann Arbor, MI 48109

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NE
801 North Randolph Street, Rm 732
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

F49620-97-1-0504

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

20010508 057

13. ABSTRACT *(Maximum 200 words)*

The objective of this project carried out under the ULTRA program of DARPA, was to validate the potential of ultrafast quantum-effect tunneling devices for the design of ultra-dense circuits and systems. The first part of the research effort involved the design, fabrication, and characterization of RTDs and HBTs. Information about device performance was then used to design and fabricate RTD-HBT circuits for a variety of unique low power high speed digital and analog applications. To facilitate the design and characterization of RTD-based circuits, a new simulator was developed that included physics-based device models for various tunneling devices as well as convergence routines to alleviate problems associated with the negative differential resistance characteristics of RTDs. The report also presents circuit and system design activities that sought to co-integrate RTDs and MOS devices to develop a viable circuit technology for the post-shrinking VLSI era.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Project Title: NDR Devices and Their Digital Applications

Contract Number: F49620-97-1-0504

Organization: The University of Michigan

Start Date: 7/1/1997

Co-PIs: Prof. Pinaki Mazumder, Prof. George I. Haddad and Dr. Jack R. East

Affiliation: Department of Electrical Engineering and Computer Science
The University of Michigan

Address: 2215 EECS, 1301 Beal Avenue
Ann Arbor, MI 48109-2122

Telephone: (734) 763-2107

1 Summary

The objective of this project carried out under the ULTRA program of DARPA, was to validate the potential of ultrafast quantum-effect tunneling devices for the design of ultra-dense circuits and systems. The results of the research are presented in this report. The first part of the research effort involved the design, fabrication and characterization of RTDs and HBTs. Information about device performance was then used to design and fabricate RTD-HBT circuits for a variety of unique low power high speed digital and analog applications. To facilitate the design and characterization of RTD-based circuits, a new simulator was developed that included physics-based device models for various tunneling devices as well convergence routines to alleviate problems associated with the negative differential resistance characteristics of RTDs. The report also presents circuit and system design activities that sought to co-integrate RTDs and MOS devices to develop a viable circuit technology for the post-shrinking VLSI era.

2 Device Design and Characterization

InP based RTDs and HBTs were the building blocks for the circuits developed in this project. Several material sources were available; The Solid State Electronics Laboratory at The University of Michigan, Hughes Research Laboratories and Ovation, a commercial vendor. Material quality was an important issue in the research. Nominally similar designs from different sources produced different device characteristics. Most of the final circuits were designed using information from calibration devices on the same wafer. Even monolayer thickness changes in the RTD specifications can greatly modify the device properties.

The first devices were RTDs. The initial structures consisted of AlAs barriers and InGaAs quantum wells. The important device parameters are the peak current density, the peak voltage and the peak to valley current ratio. The peak current density along with the device capacitance determines the switching speed and the peak voltage is a factor in the circuit DC power consumption. Peak to valley current ratios greater than 5 to 10 are needed in order to design reasonable digital functions. The initial wafer designs were used to investigate the tradeoffs in layer design. Barrier widths between 20 and 42 angstroms were used. The final peak current densities were finally optimized to be on the order of 15 KA/cm^2. The experimental room temperature PVR was 25.9. Lower current densities limited the switching speed and current density above

105 A/cm^2 were difficult to match to the experimental transistors and had a high peak voltage due to voltage drop in the parasitic resistance. However, these devices still had a peak voltage greater than 0.5 volts. This higher voltage would increase the power consumption in the final circuits.

The next step was to design RTDs with deeper InAs wells within the RTD quantum well. The lattice mismatched well is more difficult to grow using MBE, but this structure produced the lower peak voltages needed for low power operation. The deep well devices have a peak voltage of 0.25 volts, a PVR of 10.2 and a peak current density of 6.8 KA/cm^2. These devices were used in the digital circuit designs.

Transistors are needed in the digital circuits to provide current sources and input-output isolation. The main requirements are high collector breakdown voltages to provide a large logic swing, high f_t and f_{max} to match the speed of the RTDs and a process technology that will allow cointegration of the HBTs and RTDs on a single wafer. Although a variety of single and double HBT designs are available, the best match to the RTD technology was an InAlAs emitter, InGaAs base and collector structure. The emitter and base profile were standard, but several collector designs were investigated. The tradeoff is between the frequency response and the collector breakdown voltage. The combination of the collector width and the collector n-type doping can be varied to obtain the desired operating voltage. Designs with breakdown voltages of approximately 3 volts and approximately 5 volts were fabricated and tested. The DC current gain was 195 for the low breakdown voltage device and 100 for the higher voltage structure. The low voltage device had an f_t of 59 GHz and an f_{max} of 87 GHz. The high voltage device had corresponding values of 53 and 35 GHz.

The combination of the RTD on top of the emitter of the HBT increases the total height of the fabricated structure. This makes the fabrication more difficult. Step coverage in particular can be a problem. A new fabrication process using air bridge connections to bonding pads was developed to overcome this problem. This process had a much better yield than the more conventional via hole contact process that had been used earlier. However the parasitic capacitance was slightly larger.

The next step in the research was device characterization to obtain equivalent circuit models for circuit design. Cold device measurements were used to extract the parasitic inductances and resistances. TLM patterns were also used to measure resistances. Small signal bias dependent S parameter measurements from 2 to 26 GHz were used to obtain high frequency data. This data was fitted to a small signal device model using LIBRA.

Optical devices based on RTD-HBT circuits were also investigated. PIN optical photodetectors were fabricated using the base-collector-subcollector portion of the HBT. We also investigated the design and fabrication of silicon based conventional tunnel diodes as an alternative NDR device. The combination of silicon based NDR devices with existing CMOS technology would allow a variety of new QMOS (Quantum MOS) circuits to be developed. They could combine many of the low power and self-latching properties of the higher speed III-V circuits discussed elsewhere in this report with the low cost of a conventional silicon fabrication. The major design requirement for conventional or Esaki tunnel diode operation is a very narrow zero bias depletion layer and degenerate doping on the N and P sides of a PN junction. Abrupt doping changes with doping levels above 10^{20}/cm^3 and depletion layer widths smaller than 60 to 80 angstroms are needed. Several fabrication approaches including diffused junctions and silicon MBE layers grown at The Naval Research Laboratory and Hughes Research Laboratory were investigated. Silicon MBE PIN structures with a very thin I layer to prevent interdiffusion during MBE growth had negative differential resistance at room temperature. However the peak to valley current ratio was small, about 1.05. Tunnel diodes with a Ge spacer layer to produce a lower bandgap and corresponding smaller tunnel barrier had NDR with a peak to valley current ratio of 1.16, a peak current density of 3.12 KA/cm^2 and a peak voltage of 0.62 volts. This data shows the potential for silicon based NDR devices, but the performance, in particular

2

the peak to valley ratio, will have to be improved in order to be useful in circuit applications.

3 Circuit Fabrication and Measurement

The next step in the research was design of RTD-HBT circuits. The device information obtained from experimental measurements was used in the circuit simulators discussed elsewhere in this report to predict the performance of several sirʔɹe digital circuits and an optical receiver. The circuits were then fabricated and characterized. The circuits that were fabricated included a simple inverter, a C-element, an inverted majority gate, a MOBILE gate, ᵃ photoreceiver and a ring oscillator. A brief summary of these circuits will be given here. Additional details will be available in a Ph.D. thesis by Cheng Hui Lin. Logic functions were measured on wafer using a PicoProbe probe card and a 10 Gb/sec Anritsu pattern generator and a 50 GHz Tektronics digital sampling ocsilloscope. The measurements were limited by the speed of the pattern generator.

The first circuit was a static inverter, a combination of an RTD and an HBT. This simple circuit was tested up to 10 GHz using the pattern generator. The inverter was biased at Vcc = 2.5 volts and dissipated 6 mW. The output voltage swing was 110 mV. This was less than the predicted swing of 500 mV due to loading of the inverter by the 50 W measurement system. Future circuit designs will need to include output circuits to buffer the digital circuits from the measurement system. The circuit photomicrograph and measurement traces are illustrated in Fig. 1.
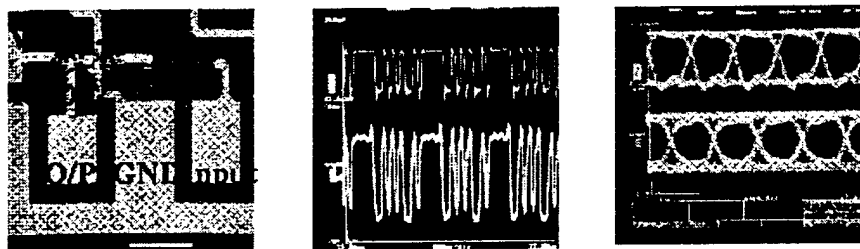

Figure 1: RTD-HBT inverter photomicrograph and measurement traces.

The next circuit was a 3 stage ring oscillator. This circuit allows direct measurement of the speed of the inverters without the frequency limitations of the 10 GHz pattern generator. The circuit was fabricated with 2 x 5 m emitter HBts and 2 x 5 m RTDs. The ring oscillator operated with a frequency of 17.32 GHz. The estimated propagation delay for the interconnects was 1.33 ps and the resulting inverter delay was 18.79 ps. The corresponding frequency is 53 GHz. Since the measured f_t of these transistors is 55 GHz, this circuit is operating operating very close the the limit of the transistors. The RTD loads allow this fast operation. The three stage oscillator consumed 2.3 mW from a 1.3 volt supply. The phase noise of these ring oscillators was also measured. The phase noise was 88.8 dBc/Hz 1 MHz away from the carrier, approximately 10 dB worse than CMOS based ring oscillators at lower frequencies.This is to be expected with the nonlinear switching of the RTD loads in these circuits. The ring oscillator fabrication results are illustrated in Fig. 2

Several other circuits were fabricated and tested. However their more complex input-output functions could not be tested at GHz frequencies with our existing equipment. The circuit performance was measured at lower frequencies and simulations were used to estimate the high frequency limitations.

The next circuit tested was a minority gate. This circuit was fabricated with 5 x 5 m emitter HBTs, twice
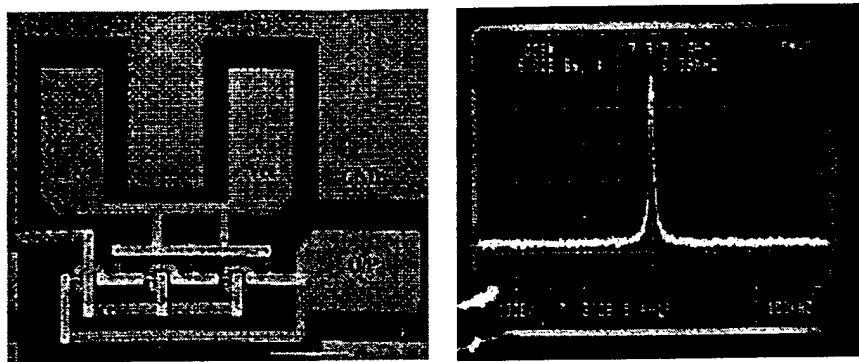
Figure 2: RTD-HBT 3-stage ring oscillator photomicrograph and measurement traces.

the emitter size of the ring oscillator. This circuit was tested at 1.6 MHz for the proper input-output function. The circuit performed well with a noise margin of 1.68 volts. We were unable to test this circuit at higher frequencies because of its pad configuration and limitations on the pattern generator. However, the higher frequency performance was investigated using a computer simulation with device models obtained from earlier device measurements. The minority gate can operate up the 10 GHz with a noise margin greater than 0.8 volts. The switching speed is mainly limited by capacitive feedthrough on the rising and falling edges of the input pulses causing false output values. The frequency response can be improved by refabricating the circuit with smaller emitters.

The final circuit to be discussed is an RTD-HBT based photoreceiver. This circuit is an RTD-HBT MOBILE inverting gate with a PIN photodetector formed in the base-collector-subcollector layers of the transistor. The RTD acts as a load that is switched with very small changes in the PIN photocurrent. The circuit includes a single HBT, 2 RTDs and a PIN diode. The PIN diode was 50 m in diameter and had an optical response of 0.7 A/W. MOBILE operation depends on a small difference in the peak current between 2 RTDs with slightly different areas. In this circuit the RTDs were 2x5 and 2x6 m square and the peak current difference was 190 mA. The resulting optical power needed to switch the gate is 271 mW. The circuit was tested with a 1.55 m laser modulated at 900 KHz. The circuit worked well at this low frequency with an output voltage swing of 0.74 volts and a conversion gain of 3000 V/W. The complete circuit consumed 0.21 mW from a 1.6 volt power supply. This is a very low power circuit. A conventional transimpedance amplifier would use approximately 120 mW. The input signal sensitivity for switching is modest, about 6 dBm. The sensitivity is determined by the peak current difference between the 2 RTDs in the circuit, which are determined by the area difference. If the area difference could be better controlled to produce a current difference of 10 mA, the sensitivity could be improved to 18.5 dBm.

4 CAD Tool Design for Quantum-Electronic Circuits

Under this task, the actual current-voltage, capacitance-voltage, and inductance-voltage equations of the quantum devices were incorporated into the SPICE device library. We have added piece-wise linear models of RTDs, resonant tunneling transistors (RTTs), resonant hot-electron transistors (RHETs), among others. Very recently, we have successfully added two different analytical models of RTDs which are loosely physics-based and a model of the surface tunneling transistor (STT) for which we had to make substantial changes to the convergence routines. Our previous experiences with quantum-effect devices has prepared us

4

to anticipate a variety of convergence problems when simulating these types of devices. Numerical overflow (NaN — not a number) and *time-step too small* problems are routinely encountered while simulating these devices and will invariably accompany newer models because of the inherent problems of NDR characteristics. We studied these problems by means of extensive simulation of a large number of benchmark NDR circuits.

After incorporating the physics-based analytical models of the RTD into SPICE3f5, we noticed that several different types of problems appeared to occur for a variety of circuits. These problems could arise during DC simulation as also during transient response simulation of digital quan⁺.m-effect circuits. We carefully documented all these problems and have studied them in great detail. Most of these problems are quite intricate in nature and are also very challenging. Below we present a case in point.

Let us consider a hypothetical circuit, shown in Fig. 3(a) consisting of two RTDs, D1 and D2, connected in series between V_{cc} and ground such that D1 is the load and D2 the driver. The I-V characteristics of the two diodes are such that D1 has a higher peak current(I_p) than D2. When V_{cc} is ramped up from 0V, the output should switch from 0 to 1 since D1 has a higher I_p.

However, for several different combinations of I-V curves of D1 and D2, with $I_{p1} > I_{p2}$, it has been observed that the simulation results do not show an output switch for a circuit that should switch. Here, we show what causes this problem. The circuit can be described by the simple nodal equation:

$$f(v_0) = i_1(v_{cc} - v_0) - i_2(v_0) = 0. \tag{4.1}$$

Fig. 3(b) shows $f(v_0)$ for various values of V_{cc} (V_{cc} is being ramped up from 0 to V). If the time-step is not small enough, situation c will be skipped and we may have $a \to b \to d$. In that case, the starting seed for d will be the solution of b which will be closer to root no. 2 in Fig. 3(b), situation d, which is not the right DC solution. The root no. 3 is the right DC solution for this case.
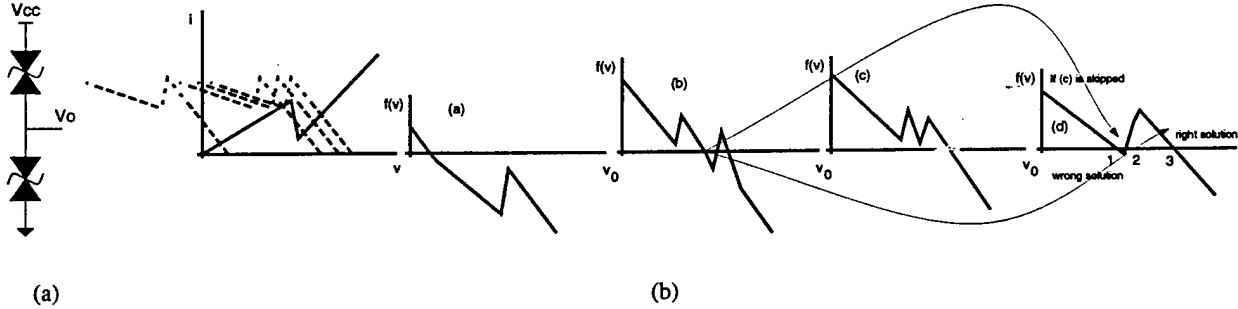

Figure 3: (a) Series connected RTD circuit, (b) the coarse time step problem.

So, what we observe is that the sudden increase in the number of solutions for the nodal equations can cause false convergence if the solutions are closely spaced and the time-step is not small enough to detect the change. This problem can actually be traced to the NDR of the quantum devices which gives rise to the abruptness of the change in the number of solutions.

SPICE chooses its time-steps by means of LTE (Local Truncation Error) method. Based on two consecutive time-point solutions, it tries to predict the solution at the current time point by means of quadratic extrapolation. The error it calculates is the error of this prediction from the true solution. If this error is

5

large, the time-step is reduced. Otherwise, it is *increased*. This type of adaptive time-step choice is done to achieve a tradeoff between simulation speed and accuracy. Herein lies the main problem so far as the above hypothetical circuit simulation is concerned. Situation *c* may totally get bypassed as a result. This is the essence of, what we call, the *coarse time-step problem*.

Thus, in this task, we first derived usable analytical large-signal and small-signal models for the RTD based on the model derived under the previous task. These models were then added to the SPICE3f5 device library. Subsequently, we performed extensive simulation of realistic medium and large-scale quantum-effect circuits, and studied DC and transient convergence issues. Based on these studies, we derived different problem syndromes for which we developed algorithmic solution techniques along the lines of our previous works in this area.

5 Quantum MOS Circuit and System Design

In this part of the project we developed novel digital and multivalued logic families using co-integrated resonant tunneling diodes (RTDs) and conventional MOS transistors which will extend the frontier of Si-Based device technologies well into the post-shrinking era, with improved circuit performance, reduced device count while maintaining very high packing densities.

The major accomplishments in this area are listed below.

- **Development of QMOS logic families**

    - Static QMOS
    - Self-latching bistable QMOS
    - Pseudo-bistable QMOS
    - Threshold-mode QMOS

- **Development of QMOS flip-flop circuits**

    - D, S-R and T edge-triggered TSPC flip-flops
    - Compact master-slave bistable and pseudo-bistable flip-flops

- **QMOS multi-valued logic circuits**

    - Signed-digit adder circuit for elimination of carry propagation
    - Parallel multiplier with multi-valued SDFA

- **Theoretical analysis and performance projection of QMOS**

    - Speed, power and noise margin analysis
    - Clocking and slew rate issues
    - RTD and MOSFET device matching
    - Statistical simulation studies

- **Gate-level pipelining scheme for QMOS**

    - Elimination of delay, area and power overhead of discrete latches

- Maximization of pipelined system throughput

- Efficient implementation of deeply pipelined communication systems

- **System design using QMOS circuits**

  - 32-bit parallel correlator

  - 32-bit direct-digital frequency synthesizer

  - 4-bit turbo-code decoder optimization

  - Pipelined carry-save multiplier

  - System prototype implementation and fabrication using generic CMOS process

5.1 QMOS logic families

A set of QMOS digital logic families have been developed for application-specific use. Combinational QMOS logic yields the simplest and most robust logic circuits of any of the QMOS logic family. It consists of an RTD pull-up load and a pull-down network of n-transistors that determines the logic operation performed by the gate. The possibility of two stable RTD operating points at the same current level, due to NDR characteristics, makes bistable QMOS logic feasible. We have developed a bistable QMOS logic family that offers the following advantages.

- Reduced device count as compared to CMOS

- Lower power delay product as compared to CMOS

- Alleviates charge sharing and coupling problems of dynamic CMOS

- Gate level pipelining without latch area/delay overhead

- Static as well as dynamic logic implementations

- Pseudo-bistable and bistable-mode operation possible

Fig. 4(a) shows a circuit diagram of a QMOS bistable gate, while Fig. 4(b) shows the load lines of the circuit that explain its operation principle. Such bistable logic elements are inherently self-latching, providing a compact topology for efficient implementation of pipelined circuits. A latched QMOS logic topology has been developed that eases some device matching constraints between the RTDs and MOS transistors. Other logic topologies are derived from these basic circuits. Pseudobistable QMOS logic elements have been demonstrated that require an evaluation pulse only for one transition direction of the output, i.e. either high-to-low or low-to-high. Threshold-mode QMOS logic circuits have been designed along with extension to implementation of weighted threshold logic. Functional completeness of the logic families is demonstrated via design of NAND and NOR logic structures. Design considerations for each logic family have been quantified along with state transition sequences for circuit operation. These yield systematic design procedures for QMOS logic topologies along with understanding of possible failure modes of QMOS circuits.
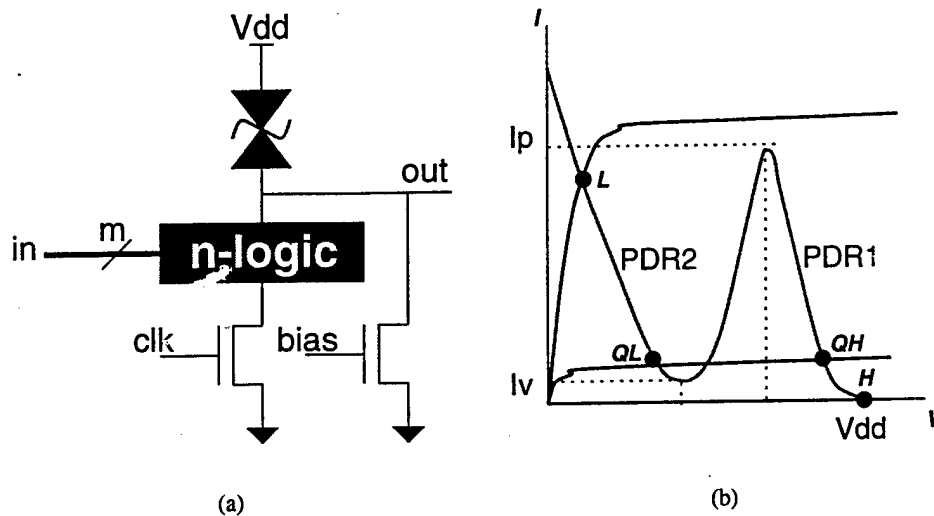
Figure 4: Bistable-mode QMOS logic gate (a) schematic, (b) load lines for circuit operation.

5.2 QMOS flip-flop circuits

We have designed of a family of edge-triggered flip-flops using RTDs and MOSFETs. A Monte-Carlo simulation of the QMOS D flip-flop and a conventional true single phase clock (TSPC) CMOS flip-flop using the same MOS devices shows that the QMOS flip-flop operates at a higher frequency than the TSPC flip-flop. Table 1 shows the comparison between a QMOS D flip-flop and a TSPC D flip-flop implemented in 0.35 micron CMOS technology. For the normalized area comparison, it is assumed that the RTDs can be vertically integrated on top of the source and drain regions of MOS devices and hence do not contribute to silicon area.

Table 1: Comparison of QMOS D Flip-flop with TSPC CMOS D Flip-flop

Parameter	CMOS	QMOS
Area (normalized)	1	0.75
Setup time (*ns*)	0.1	0.06
Hold time (*ns*)	0.2	0.09
Rise time (*ns*)	0.2	0.09
Fall time (*ns*)	0.12	0.08
Maximum speed (*GHz*)	2.5	6.6
Power (*μW*)	129	34
Power-delay (*fJ*)	81	17

The D flip-flop can be easily modified to operate as an S-R flip-flop or a T flip-flop. The T flip-flop uses an XOR gate to generate the controlling feedback from the flip-flop outputs. An XOR gate can be implemented using just two RTDs and two n-type MOSFETs and hence we achieve an extremely compact implementation of a T flip-flop. Negative edge-triggered flip-flops can be derived from their positive edge-triggered version by interchanging the two clock transistors. Asynchronous set/reset operations can also be

8

accomplished by addition of two transistors to the output node. Bistable QMOS gates can also be cascaded to obtain master-slave QMOS flip-flop circuits.

5.3 Theoretical analysis and performance projection of QMOS

In order to understand the behavior and potential advantages of QMOS circuits in detail, it is essential to analyze and characterize these circuits. Theoretical analysis of QMOS circuits yielded analytical expressions for noise margin, switching delay and power dissipation. These results were used to analytically compare QMOS and CMOS circuit topologies. Circuit network topology comparisons were used to contrast QMOS and CMOS flip-flop circuits in order to assess their performance. QMOS circuits were also characterized using simulation studies. These studies compared QMOS circuits with their closest CMOS counterparts to illustrate advantages and disadvantages of QMOS circuits. The effects of technology scaling on QMOS circuits were investigated to determine the viability of these circuits in the deep submicron VLSI era.

While analysis expressions are presented in attached publications, here we discuss the general trend of circuit behavior while comparing QMOS and CMOS inverters in terms of a figure of merit that is a function of the circuit speed, power, noise margin and area. Fig. 5(a) shows the comparison of the QMOS noise margin with the CMOS noise margin as a function of RTD peak current, I_p and transistor gain factor, β_n. For all normal operating conditions, the QMOS noise margin is superior to the CMOS noise margin. As the I_p increases, it takes a larger input swing to switch the RTD from $PDR1$ to $PDR2$ and that is reflected in the improved noise margin at higher I_p values. Also, as β_n increases, the influence of the NMOS transistor in determining noise margin increases and hence the ratio of the QMOS and CMOS noise margin reduces. This trend is also validated in Fig. 5(a).

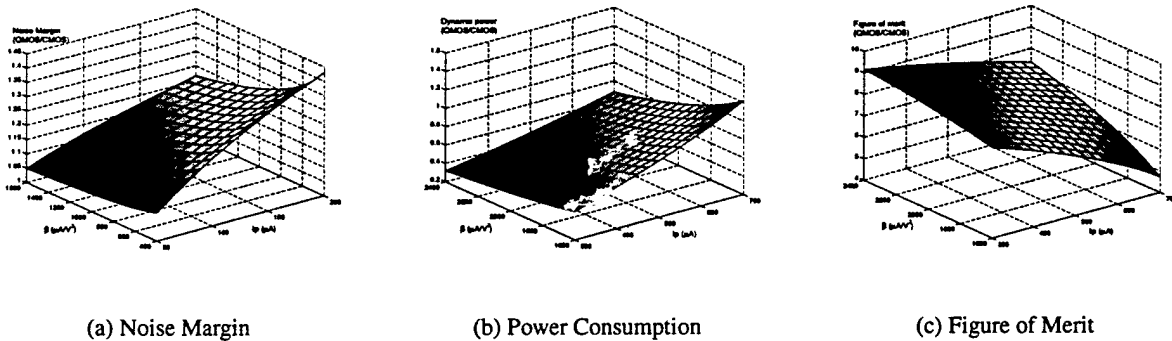| (a) Noise Margin | (b) Power Consumption | (c) Figure of Merit |

Figure 5: Analytical comparison of QMOS and CMOS.

Similar studies will be conducted for the rise and fall times, and power dissipation of digital building blocks. Fig. 5(b) shows preliminary results of comparison between QMOS and CMOS power consumption ratios. A figure of merit for the inverter design comparison is defined as follows.

$$F_{QMOS/CMOS} = \frac{Noise\ Margin\ Ratio}{PowerDelay\ Ratio \times Area\ Ratio}. \tag{5.1}$$

Since RTDs can provide higher current density than PMOS devices, their size is smaller than a PMOS transistor used in an inverter. Also, RTDs can potentially be vertically integrated on top of source/drain

regions of FETs. This leads to further possible area reduction. The figure of merit comparison for QMOS and CMOS inverters is illustrated in Fig. 5(c) for varying values of β_n and I_p.

Comparison between simulation and analysis results were performed to validate the analysis techniques. A sample result of the theoretical and simulation-based delay times plotted in Fig. 6 shows a good approximation between simulation and theoretical results.
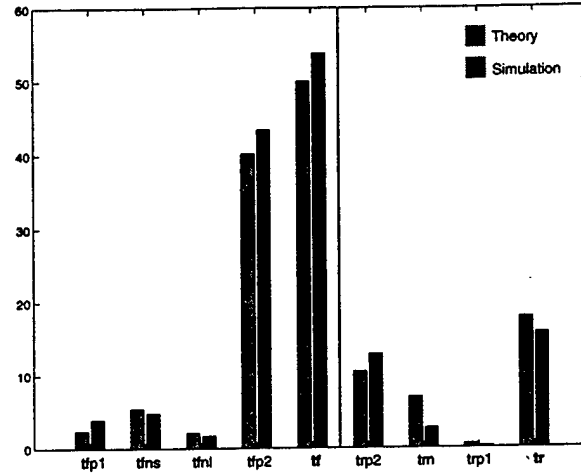

Figure 6: Comparison of theoretical and simulation-based delay times of the QMOS bistable inverter.

The accuracy of the analytical expressions for power consumption is also evaluated with the help of SPICE simulation. The results of the energy consumption of the bistable inverters are shown in Table 2 along with theoretical and simulated power consumption figures. The simulated and theoretical values are within 10 % of each other, indicating a good match.

Table 2: Comparison of analysis and simulation results for power consumption.

	E_{eval1} (fJ)	E_{tfp1} (fJ)	E_{tfn} (fJ)	E_{tfp2} (fJ)	E_{eval0} (fJ)	E_{l1} (fJ)	E_{l0} (fJ)	P_{theory} (μW)	$P_{simulated}$ (μW)
Example A	1.85	0.19	0.86	3.12	4.17	3.75	3.93	68.4	72.8
Example B	2.66	0.31	1.42	3.47	5.2	4.55	4.70	70.5	75.6

Studies were conducted to determine optimum RTD parameters in order to maximize QMOS circuit performance. Again theoretical analysis as well as simulation studies were used to conduct these experiments and for a sample CMOS process, a good match between the results of optimized RTD characteristics predicted by analysis and those obtained via optimization software is observed, and this is graphically illustrated in Fig. 7.

Simulation-based comparison of various CMOS and QMOS circuits was performed to illustrate the potential benefits of QMOS circuits. Table 3 shows the comparison between the CMOS and QMOS 31-stage ring oscillators. The larger number of stages allows rail-to-rail switching of stage outputs. The simulation results show a twofold improvement in speed for the QMOS ring oscillator over the CMOS version. A more representative indication of the speedup of QMOS is the difference between the rise and fall times of the two circuits.
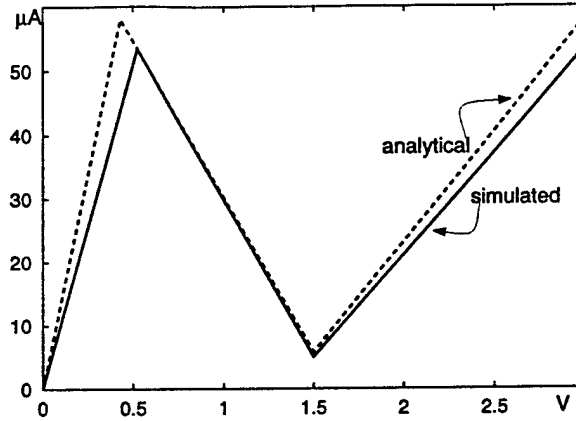
10

Figure 7: Comparison of optimized RTD salient points using analysis and simulation.

Table 3: Comparison of QMOS and CMOS 31-stage ring oscillators.

Parameter	CMOS	QMOS
Period (ns)	8.54	4.1
t_r (ps)	400	130
t_f (ps)	180	140

Table 4 shows the comparison between a dynamic CMOS inverter and a bistable QMOS inverter. It indicates a substantial improvement in switching speed for the QMOS bistable inverter over the dynamic CMOS inverter compensating for the higher absolute power of the QMOS inverter due to its higher operating frequency. Also, although the QMOS inverter has only one fewer active device than the dynamic CMOS inverter, it shows threefold improvement in area due to the absence of PMOS transistors that have to be sized three times as large as NMOS transistors.

A QMOS master-slave flip-flop consisting of two cascaded bistable inverters is compared with a C^2MOS flip-flop that offers a robust and race-free implementation of a dynamic CMOS master-slave flip-flop. Results of the comparison are tabulated in Table 5. Since the flip-flop is a clocked cascaded circuit, the minimum width of the clock pulse is not determined by the average propagation delay, but by the worse case propagation delay.

5.4 QMOS gate-level pipelining

The self-latching nature of basic QMOS gates, and the ability to implement extremely compact gates using QMOS logic leads to new possibilities for fast adder designs that utilize gate-level pipelining to provide very high addition throughput. Thus, a high-throughput system design technique is made possible in which primitive logic gates also perform the latching function without the necessity for external latches. This eliminates delay area, and power overhead in pipelined systems, thus further improving speed and throughput of deeply pipelined systems over what can be gained as a result of the picosecond switching speeds of RTDs. Using QMOS logic, a bistable full adder was designed as a two-stage logic block. For correct operation of the bistable logic gates, bias and clock pulses are required as mentioned previously. However, when

Table 4: Comparison of QMOS and CMOS clocked inverters.

Parameter	CMOS	QMOS
t_{plh} (ps)	300	73
t_{phl} (ps)	119	81
t_r (ps)	250	94
t_f (ps)	194	100
Power (μW)	22	35.8
Power-delay (fJ)	6.6	2.9
Device count	5	4
Area (normalized)	3	1
Area-power-delay (normalized)	6.8	1

Table 5: Comparison of QMOS and CMOS master-slave flip-flops.

Parameter	CMOS	QMOS
Period (ps)	636	214
Power (μW)	32	61
Power-delay (fJ)	20.4	13.1
Device count	8	8
Area (normalized)	1.7	1
Area-power-delay (normalized)	2.7	1

multiple gates are cascaded, as is the case in a full adder, a gate must be clocked only after all its inputs have been correctly evaluated. This requires a multiphase clocking scheme in which each gate is evaluated in a different phase than its fanins and fanouts. For cascaded QMOS gates, a two-phase clocking scheme is used. Fig. 8 shows the schematic of the full adder circuit. The majority function that gives the carry output is evaluated on *Phase*1 of the clock. Bistable buffers clocked on *Phase*1 are required to synchronize inputs to the sum stage of the circuit which is evaluated on *Phase*2 of the clock. The two-phase clock operates at 2 GHz and the average propagation delay of the adder is 220 ps.
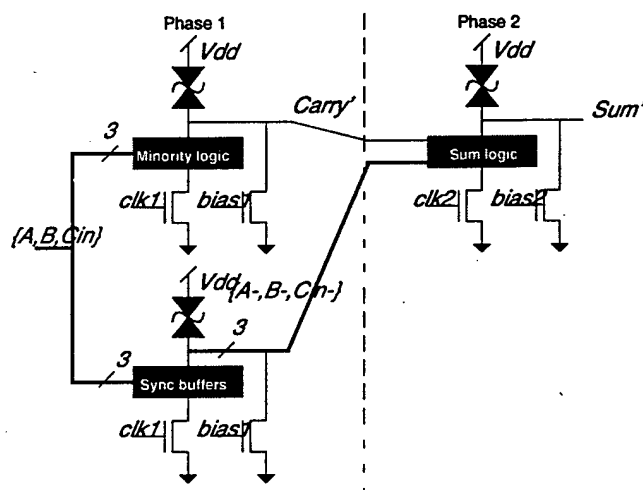

Figure 8: Pipelined QMOS full adder schematic.

In this circuit, two computations can be active concurrently and this is how fine-grained pipelining improves the throughput of the system. The QMOS bistable full adder uses 5 RTDs and 20 NMOS transistors. To convert a standard 24-MOSFET static CMOS adder to a similar gate-level pipelined adder, one would require an additional 40 transistors for 5 latches required for the carry, the sum and the three input signals. The addition of these latches would provide the pipelining advantage in standard CMOS but would increase the stage delay and area due to the additional latches. Thus, the QMOS logic family has advantage over conventional CMOS logic in terms of greater circuit compactness and improved speed. Table 6 shows a comparison of a QMOS bistable full adder and a clocked dynamic NORA CMOS full adder each designed using 0.25 micron MOS devices. For this comparison, the circuits were operated at the CMOS process recommended 2.5 V, so as to optimize the CMOS adder design. Also, the power dissipation of the QMOS and CMOS circuits were kept close to each other in order to measure the effect on circuit delay. While the QMOS circuit requires only 7 fewer devices than the NORA CMOS adder, due to the large size of the PMOS devices as compared to NMOS transistors, the area of the CMOS circuit is more than twice that of the QMOS circuit.

The use of gate-level pipelined 1-bit adders in designing larger computing subsystems with high throughput, such as multipliers, high-speed parallel correlators and turbo-decoders is discussed in the following sections. Also, the concept of self-timed QMOS pipelined datapaths is introduced to allow design of systems with reduced clocking complexity.

13

Table 6: Comparison of QMOS pipelined adder with NORA CMOS clocked dynamic adder.

Pipelined Adder	NORA CMOS	QMOS
Device Count	47	40
Normalized Area	2.14	1
Power (mW)	1.03	0.96
Maximum Delay (ps)	350	181
Power-Delay (pJ)	0.36	0.17

5.5 QMOS system design

Multiplication is important function in arithmetic logic units (ALUs) of microprocessors, and also is a critical function in signal processing chips. It often forms a critical path operation and hence many techniques such as Booth recoding and Wallace tree are used to speed up multiplication. These implementations lead to $O(log_2 N)$ multiplication time for N-bit operands. These implementations, however, result in irregular layout, and circuit design that is not modular, causing increased design turnaround time and cost. On the other hand, carry save multipliers provide a regular design and layout strategy but have an $O(N)$ multiplication time making them slow. In order to combine the ease of design and layout along with improved multiplication time, pipelining of multiplier designs is a common practice in VLSI design of arithmetic units. However, pipelining conventional multiplier circuit has some disadvantages. Adding static latches to basic multiplier cell outputs can lead to as much as 50% delay overhead for the latch operation. This critically limits the throughput of the multiplier. We propose the design of a pipelined carry-save multiplier using QMOS bistable gates that achieves *constant-time* multiplication at the expense of latency caused by pipelining. Such a multiplier is useful in signal-processing chips where multiplication is a critical operation and latency is not important.

Using QMOS bistable logic, a pipelined multiplier cell can be constructed as shown in Fig. 9 that forms the basis for constant-time multiplication.
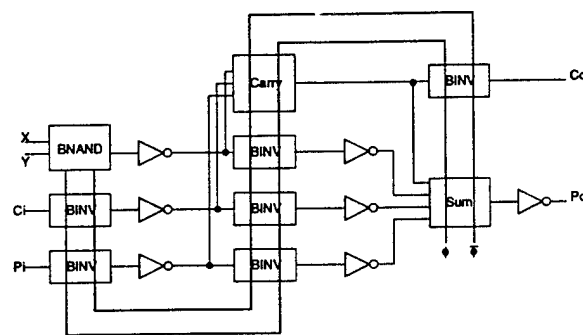

Figure 9: QMOS pipelined multiplier cell.

Using the pipelined multiplier cell of Fig. 9, the parallel multiplier can be pipelined at the gate level with small modifications that do not compromise layout regularity and modular design philosophy of the carry-save multiplier. The block diagram of a QMOS pipelined 4x4 multiplier is shown in Fig. 10.

The core layout consisting of pipelined multiplier cells can be modularly designed as in the case of
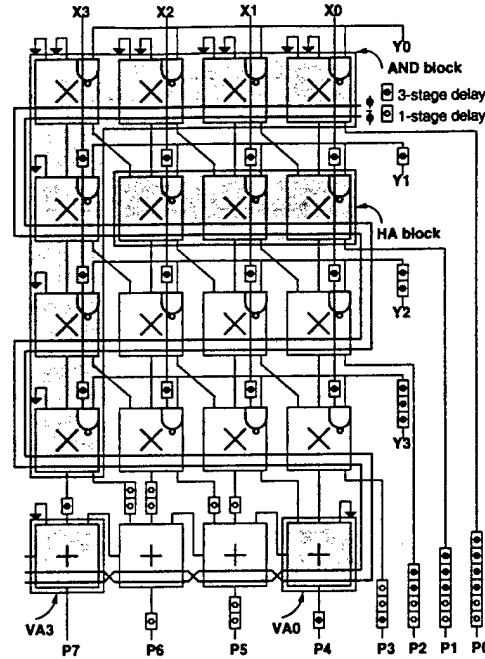
14

Figure 10: QMOS 4×4 pipelined multiplier schematic.

a conventional carry-save multiplier. Delay stages are introduced to synchronize the multiplier X and Y inputs to different multiplier rows. These delay elements for the X signals can be easily incorporated into the multiplier cell design in order to retain modularity. The vector-merging adders required to add the final carry outputs of the multiplier cells need successive multiplier column outputs to be increasingly delayed to maintain overall synchronization of the design. Since the multiplier outputs are generated at different stages, bistable buffers are required to synchronize the outputs. As in a conventional carry-save multiplier, optimizations can be made to reduce the circuit complexity.

Binary parallel concatenated recursive systematic convolutional codes, termed turbo-codes have exceptional error correcting capabilities that are vital for good signal reception quality in mobile wireless communication systems. The decoding complexity of these algorithms makes their VLSI implementation difficult. The conventional CMOS implementation of such a scheme results in poor data rates due to the large amount of combinational logic utilized in such an implementation. Since this turbo-code decoder has no data dependence, it is an ideal candidate for gate-level pipelining in order to improve the throughput and data rate.

Fig. 11 illustrates the block diagram of a 4-bit slice turbo-code decoder. The core of this 4-bit slice is formed by two identical 4-bit component decoders. The first decoder operates on the input data stream along with redundant information and a-priori information. The second decoder uses the delayed and interleaved input stream along with the interleaved output of the first decoder and a second set of redundant information to generate an output that is fed to a decision circuit that generates the decoded output.

The comparison of the QMOS turbo-code decoder with fully pipelined static and dynamic CMOS implementations is presented in Table 7.
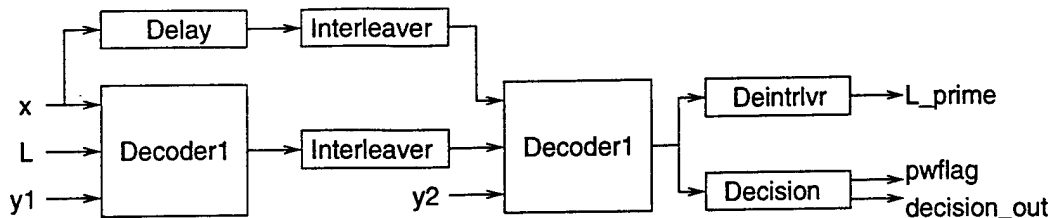
15

Figure 11: 4-bit turbo-code decoder

Table 7: 4-bit Turbo-code Decoder Performance Comparison

Turbo-code Decoder	Static CMOS	Dynamic CMOS	QMOS
Device Count	51082	29974	16652
Power (mW)	140	147	139
Speed (GHz)	1.0	1.7	3.3
Power-Delay (pJ)	140	87	42

5.6 Fabrication of QMOS circuits

We have implemented the following QMOS logic circuits that are currently being processed at Georgia Tech University after fabrication by MOSIS.

- QMOS 2-input static NAND gate

- QMOS clocked inverter

- QMOS 2-input bistable NOR gate

- QMOS shift register circuit

- QMOS edge-triggered flip-flop circuits

- QMOS bistable half-adder circuit

The CMOS circuitry has been designed using MOSIS SCMOS 0.6 micron design rules and submitted for fabrication on an HP AMOS-14TB run on March 1, 1999. For the aforementioned circuits, a 4 sq. micron series-connected RTD pair was used in each design. The circuits will be tested by on-chip test generation circuitry. A layout plot of the chip is shown in Figure 12.

5.7 Multivalued Signed-Digit Adder Prototype

A new multiple-valued signed digit adder prototype was designed and fabricated. The circuit combines RTDs and MOS transistors in the same package. The CMOS circuits were designed for a 0.5-micron process and they were fabricated through the MOSIS system. Currently, the circuits are being processed for the attachment of the resonant-tunneling diodes. Laboratory measurements and tests will be performed in
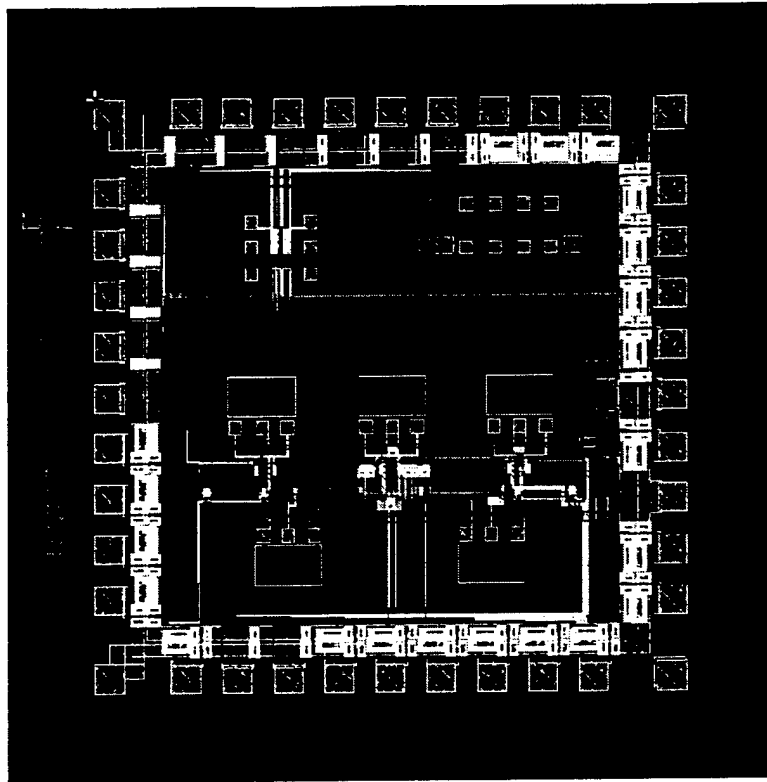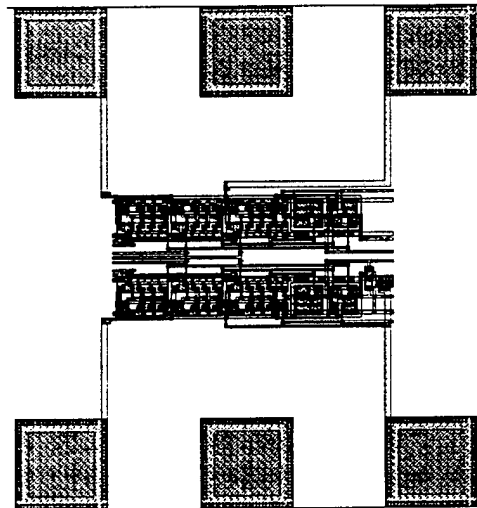
16

Figure 12: QMOS chip layout

Figure 13: CMOS layout of the signed-digit adder prototype which is being built. The layout shown is for two adder cells.

Michigan once the processing work is finished. Figure 13 shows the layout prepared for the new signed-digit adder prototype. Notice that the layout shown includes two adder circuits.

Fig. 14(a) depicts a block diagram of the signed-digit addition approach proposed in this work. Lines x_i, y_i, c_i, w_i, and s_i are three-valued, current-mode signals. Addition of x_i and y_i is achieved by simple wired-summation of currents. The function of the SDFA block is to convert the summation of input signal, z, to a two-digit representation of the sum given by digits c and w; that is, $rc + w = z$ where $r = 2$. The final sum output, s_i, is obtained by current-addition of the interim sum output, w_i, and the incoming carry signal, c_{i-1}. The transfer functions of the SDFA block are defined so that w and c always represent the arithmetic value of $x + y$. Fig. 14(b) shows the transfer functions for the interim sum, w, and the carry, c, signals in the SDFA cell. All the digits in the graph are positive because the circuit will use only positive currents. In this case, the signed-digit 0 is represented by a current level "3", digit -2 is represented by current "1", and so on. There are two pairs of transfer functions, and the working pair is selected by the value of z_{i-1}. This input signal is used to determine if $c_{i-1} \neq -1$, which indicates when the SDFA cell is allowed to generate an output $w = -1$ without causing invalid s current levels to be produced. If the input z_{i-1} to the previous digit was not considered, then it would be possible to generate $w = -1$ or $w = 1$ when $c_{i-1} = -1$ or $c_{i-1} = 1$, respectively. In these cases, the final sum result would be $s = -2$ or $s = 2$, which are invalid outputs for the selected radix.
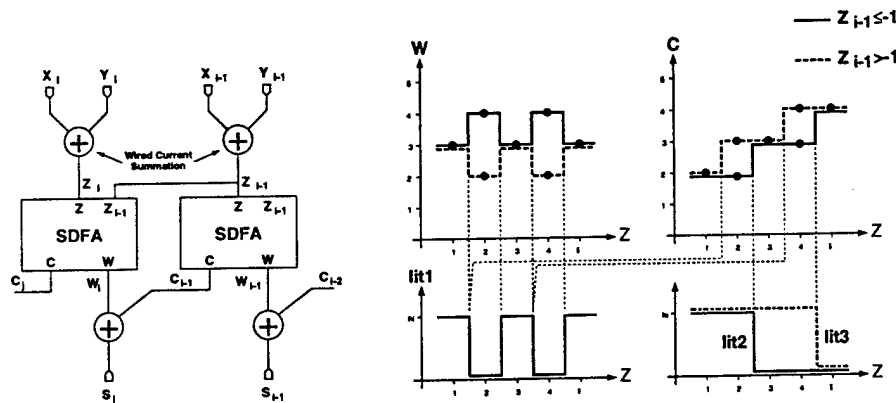

Figure 14: (a) Block digram of the implemented totally-parallel addition approach and (b) Transfer function of the SDFA block.

5.8 Summary of QMOS activities

The research in QMOS circuits and systems has yielded new logic families for compact and high-performance implementations of basic gates. These logic gates have been used to design static flip-flop circuits that do not require feedback to maintain latched outputs. As with any new circuit technology, qualitative and quantitative characterization has been performed to determine potential advantages and disadvantages of QMOS circuits as compared to conventional CMOS circuits. These analyses clearly show the effect of device characteristics on circuit behavior. This relationship is crucial in determining design margins in face of device and process parameter variations. In order to demonstrate the advantages of QMOS logic at the system level, various large-scale circuit examples have been considered that require the implementation of fine-grained pipelines. Such systems benefit significantly from improved throughput due to

the self-latching nature of QMOS. At the system level, implementation techniques for self-timed QMOS datapaths are suggested which alleviate the performance limitations imposed by handshaking circuitry in conventional systems.

The design considerations for each logic style have been addressed along with possible failure modes. The logic families have been extended to implement master-slave and true single-phase clocked flip-flop circuits. The absence of feedback in these circuits results in high-speed performance while reducing circuit size at the same time. The ease of integrating logic functions within the latches leads to efficient implementation of logic flip-flops. The first true single phase clock implementation of a flip-flop using NDR devices has been accomplished. The use of pseudobistable logic gates in flip-flop implementation has been shown to reduce the load on clock signal lines, a major factor limiting integrated circuit performance. Since the output load for QMOS circuit does not include any PMOS devices, the effective load capacitance of QMOS circuits is about one-third that of CMOS circuits. The main drawback of QMOS circuits is static power consumption. Circuit- and system-level techniques for minimizing power consumption have been explored. These include circuitry for removing bias under high output conditions, and gate-level pipelining to limit static power consumption. Suggestions for tailoring RTD characteristics for low power considerations are evident from theoretical analyses that project circuit performance under variety of device characteristics variations. It is desirable to have a low valley current for the RTD with a valley voltage that is close to the supply voltage. In such a situation, the bias transistor is locked to the RTD valley in its linear region, thus dissipating the minimum amount of power in order to maintain the output voltage. The theoretical analyses and SPICE-based characterization of QMOS circuits also yields important guidelines for circuit designers and device engineers in terms of optimizing circuit and device parameters, and matching RTD and MOS device parameters, for low area-power-delay product. A greater than twofold improvement in area-power-delay product in generally observed for QMOS circuits over their CMOS counterparts. Simulation projections for technology scaling show that this performance improvement of QMOS over CMOS will remain valid in the deep submicron regime.

The self-latched QMOS circuits have been used for efficient implementation of fine-grained pipelines that are useful in computing systems that require large volume of similar computations at each clock cycle with minimal data dependence, such as communication systems and signal processing systems. The elimination of the area, power and delay overhead of discrete latches leads to system throughput at gate operating speeds. Efficient implementation of QMOS self-timed datapaths alleviates the performance limitation of conventional self-timed circuits. Specific examples of large-scale circuits such as a multiplier, correlator and turbo-code decoder have been studied to assess the system-level benefits of QMOS logic. QMOS logic is a viable circuit alternative for boosting the performance of CMOS in the deep submicron regime.

6 Publications

1. C.-H. Lin, K. Yang, A. F. Gonzalez, J. R. East, P. Mazumder, G. I. Haddad, D. H. Chow, L. D. Warren, H. L. Dunlap, J. A. Roth, and S. Thomas III, "Fabrication and Characterization of RTD-HBT Inverter", 17th Biennial IEEE/Cornell University Conference on Advanced Concepts in high performance Devices, Ithaca, N.Y. August 7th - 9th , 2000

2. C.-H. Lin, K. Yang, J. R. East, G. I. Haddad, D. H. Chow, L. D. Warren, H. L. Dunlap, J. A. Roth, and S. Thomas III, "A Ring Oscillator Using an RTD-HBT Heterostructure", ICSMM-2000, pp. 120-122, Kyongju, Korea, September 25th -27th , 2000

3. C.-H. Lin, K. Yang, A. F. Gonzalez, J. R. East, P. Mazumder, and G. I. Haddad, "InP-Based High

Speed Digital Logic Gates Using an RTD/HBT Heterostructure" 11th IPRM, pp. 419-422, Davos, Switzerland, May 16th - 20th , 1999

4. C. H. Lin, K. Yang, M. Bhattacharya, X. Wang, X. Zhang, J. R. East, P. Mazumder, and G. I. Haddad, "Monolithically Integrated InP-Based Minority Logic Gate Using an RTD/HBT Heterostructure", 10th IPRM, pp. 419-422 Tsukuba, Japan, May 11th - 15th , 1998

5. A. F. Gonzalez, M. Bhattacharya, C.-H. Lin, P. Mazumder, J.R. East, and G.I. Haddad, "High-Speed digital Circuits Using Resonant-tunneling Diodes and Heterojunction Bipolar Transistors" GOMAC Mar, 2000

6. S. Kulkarni and P. Mazumder, "Quantum MOS circuit design," in European Conference on Circuit Theory and Design, August 1999.

7. T. Uemura and P. Mazumder, "Design and analysis of RTD-based high performance memory system," IEICE Transactions on Electronics, September 1999.

8. T. Uemura and P. Mazumder, "Quantum MOS sense amplifier," in Ninth Great Lakes VLSI Symposium, March 1999.

9. A. F. Gonzalez and P. Mazumder, "Multiple-valued signed-digit adder using NDR devices," IEEE Transactions on Computers, September 1998.

10. S. Kulkarni and P. Mazumder, "Full adder circuit design using RTDs and MOSFETs," in Proceedings of the Government Microelectronics Applications Conference, 1998.

11. N. Deb, J. Xiong, M. Bhattacharya, S. Kulkarni, and P. Mazumder, "Switching Speed and Power Consumption of Bistable Q-MOS circuits," in Third IEEE Silicon Nanoelectronic Workshop, Hawaii, June 1998.

12. A. F. Gonzalez and P. Mazumder, "Ulra-fast adder design using resonant tunneling diodes and MOSFETs," in Proceedings of the Government Microelectronics Applications Conference, 1997.

13. A. F. Gonzalez and P. Mazumder, "Compact signed-digit adder using multialued logic," in Proceedings of the Conference on Advanced research in VLSI, 1997.

14. M. Bhattacharya and P. Mazumder, "Noise margins of threshold logic gates containing resonant tunneling diodes," in Proceedings of the 1998 8th Great Lakes Symposium on VLSI, February 1998.

15. P. Mazumder, S. Kulkarni, M. Bhattacharya, J. P. Sun, and G. I. Haddad, "Digital circuit applications of resonant-tunneling devices," Proceedings of the IEEE, vol. 86, no. 4, April 1998.

16. M. Bhattacharya, S. Kulkarni, A. F. Gonzalez, and P. Mazumder, "Prototyping RTD-CMOS circuits using enhancement-mode NMOS negative differential-resistance circuits," International Symposium on Circuits and Systems, 2000.

17. A. F. Gonzalez, M. Bhattacharya, S. Kulkarni, and P. Mazumder, "Standard CMOS Implementation of a Multiple-Valued Logic Signed-Digit Adder Based on Negative-Differential-Resistance Devices," in Proceedings of the International Symposium on Multiple-Valued Logic, pp. 323-328, 2000.